ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

# A Novel Approach to Analyze Medicine Feedback Reviews Using Sentimental Analysis.

**Sasikumar AN[1]*, and Joshila Grace[2].**

PG Scholar[1], Assistant Professor[2], Department of Computer Science and Engineering, Sathyabama University, Chennai, Tamilnadu, India.

**ABSTRACT**

Twitter is a medium where people share their opinion freely. In this work, a method, which performs classification of tweet sentiments to analyze and identify the feedback of drugs, is proposed. Users' feedback can be categorized according to three polarities: Positive, Neutral and Negative. Opinion Mining is a knowledge extraction procedure, which could be very useful for manufacturing and service sector organizations. Users' perspective about products can influence a potential buyer who is willing to hear any constructive feedback from existing customers of a product. Now a days, social networking websites are widely used by millions of people who share their opinions about a wide variety of things. The feedback users share in these social networking sites can measure the efficiency of Drugs. In this research, we focus on opinions about drugs available in Twitter and classification methods, which classifies into three different polarities.

**Keywords:** Twitter, Sentiment Analysis, Drug Reviews, Opinion Mining

***Corresponding author***

ISSN: 0975-8585

## INTRODUCTION

Now a days, millions of users share their opinions on Internet and it constantly growing at a rapid pace and many organizations are using this overflow of data to pull out people's views towards their products. Social Networking Sites contain huge amount of data, which provides exceptional opportunities to get insights into customers' world. These unstructured text information can be analyzed to gather some vital information about the products merits and demerits. Social networking websites facilitate users to have a say, alter and rank the content, as well as to convey their individual opinions. Twitter is a predominately-used social networking site and in addition to it there are thousands of forums and blogs also available. These platforms are used by people to express their opinion about different topics, thus it is a valuable source of people's opinions. In our research, we focus on Drugs and their feedback or reviews expressed by twitter users.

Sentiment is an attitude, suggestion and conclusion encouraged by emotion. Our research focuses on using Twitter and other social networking platforms, for the job of sentiment analysis. Tweets about drugs are retrieved and processed in order to make them suitable for further processing. These tweets are generally semi structured and unstructured and hence needs to be processed carefully. Twitter data can be very easily retried using APIs. It supports REST API and Search API. Once, tweets are retrieved, we can perform a linguistic analysis. We then construct a Classifier to categorize the tweets according to three polarities: Positive, Negative and Neutral. We also use the open source dictionary WordNet version 2.1 as a warehouse to map each word according to the Parts-of-Speech Tags. The SentiWordNet is used to assign the sentiment scores.

## LITERATURE REVIEW

The categorization of polarity is the fundamental task in sentimental analysis Xing Fang and Justin Zhan [1] recommended a mathematical model to identify three different categories of polarity. Hu and Liu [2] research work focuses on words that describe positive, negative and neutral words from reviews. Gann et al[3] list out tokens calculated from Twitter review data. A lot of research work is carried out on Machine Learning to address the issues in Natural Language Processing. The research work of Lin et al[4] quite useful in Twitter's integration of machine learning tools. Bian et al[5] proposed an approach to analyze the effect of tweets on the efficiency of drugs. Johan Bollen et al[6] has proposed an unique research work on how Friend-Follower network of Twitter users could be vital in understanding the common behavior of twitter of users. Lot of research work has been done[7-12] on various aspects of twitter related sentimental analysis and efficiency measurements of drugs.

## MATERIALS AND METHODS

### Development Methodology

The proposed system is implemented as:

Step 1: Retrieve Semi-structured and Un-structured data from Twitter.
Step 2: Apply Stemming and Other Processing tasks to remove the noise and make the data consistent.
Step 3: Perform Sentiment analysis on the data using training data and keywords.
Step 4: Store processed data (with sentiment) in MSSQL Server database.
Step 5: Extract sentiments from SQL Server to visualization layer.
Step 6: Apply a classification algorithm to measure the efficiency

The proposed system has the following modules ;

1. Data Collection
2. Preprocessing
3. Sentiment Polarity Analysis
4. Classification
5. Evaluation Metrics

ISSN: 0975-8585

**DATA COLLECTION**

Twitter Data can be extracted online through Streaming API[1]. Here, we are going to retrieve data related to drugs. To achieve this we make use of drug data set. A drug data set is a collection of drug related data. Drug data sets are freely available on Internet. While retrieving tweets using Stream API and REST API we will filter the unwanted data.

**PREPROCESSING**

This is the most important step because the data retrieved is highly unstructured in nature. We have to remove the special characters (#,@) and emoticons. The overall efficiency of the proposed method is based on this step.

**SENTIMENT POLARITY ANALYSIS**

To measure the three different polarity of the reviews we have to make use of the WordNet 2.1 online dictionary to identify the polarity. Part-of-Speech(POS) tags must be analyzed to identify whether the tweets belong to Positive or Negative or Neutral category. Each tweet is assigned with a weight which is either 1 or 0 or -1 based on the polarity it belongs to.

**CLASSIFICATION**

Here, we build a classifier to predict the accuracy of polarity a tweet belongs to. With the help of training data a simple and accurate classifier is built. We propose a classifier based on K-Nearest Neighbour algorithm to predict the distance between unknown patterns and known patterns. The similarity measure is used classify the new patterns based on the known patterns.

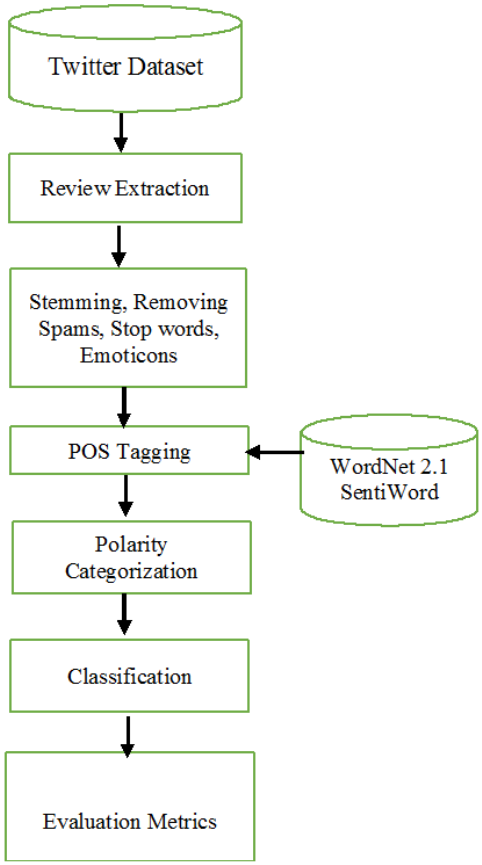


**Fig. 1. Sentimental Analysis Process**

ISSN: 0975-8585

## EVALUATION METRICS

Classification algorithms efficiency can be measured using Information Retrieval metrics. Precision can be defined as closeness of two or more measures to each other. In Machine Learning Precision can be defined as:

$$Precision = TP/(TP + FP)$$

Recall can be defined as:

$$Recall = TP/(TP + FN)$$

## EXPERIMENTAL SET UP

The proposed system is implemented using Microsoft .NET framework and MS SQL Server Database. We retrieved twitter data between January 2017 and March 2017. Nearly, 80 Drugs and considered and more than 24000 tweets were retrieved. To retrieve tweets from tweeter, the REST API is used. The first step in retrieval of tweets deal with getting access to token through oauth authentication. This is done by validating the credentials of twitter account. The next step involves removing the Spams, special characters @, # and emonticons and the words from tweets are stemmed to identify their word stem.

## RESULTS AND DISCUSSIONS

Our research work focuses on Medical Drugs data and hence we retrieve only user reviews focuses on the above. The sentimental analysis module identifies the polarity of reviews into Positive, Neutral and Negative. We then apply Sentiment Variation tracking to analyze the reviews in order to identify the merits and demerits of the drugs.The below table shows polarity of Ativan tablet:

**Table 1: Polarity of Ativan**

| Drug Name : Ativan | | |
|---|---|---|
| Positive | Negative | Neutral |
| 96 | 65 | 10 |

The medicines efficiency is influenced by its side effects. Most of the users tweets about three important side effects a medicine causes to them: Vomiting, Headache and Stomach Upset. We found that these three important factors played important role in the drug's recommendation.

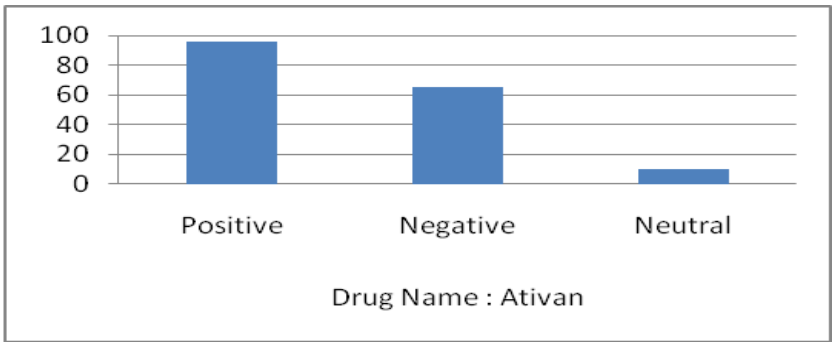


Fig. 2: Polarity of Ativan

We then used these reason candidates to further analyze the reviews. The below tablet shows Ativan tablet's efficiency is greatly affected by Vomiting.

ISSN: 0975-8585

**Table 2: Reason Candidates of Ativan**

| Drug Name : Ativan | | |
|---|---|---|
| Vomiting | Headache | Stomach Upset |
| 40 | 20 | 5 |

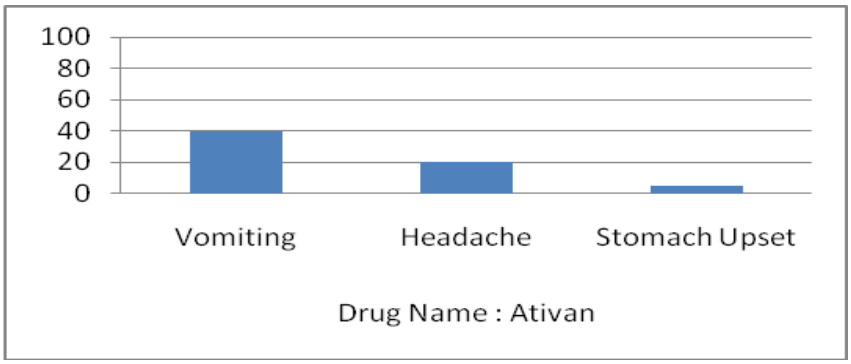


**Fig. 3: Reason Candidates of Ativan**

The evaluation metrics of classifier is measured against the following Information Retrieval metrics: Precision and Recall.

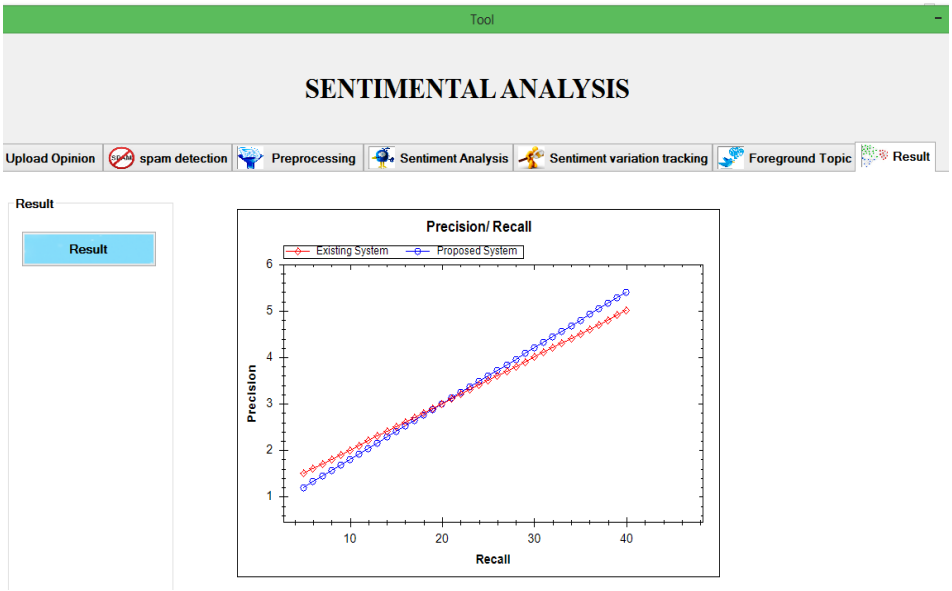


**Fig. 3. Precision and Recall Metrics**

## CONCLUSION

In this research, we implemented sentimental analysis on twitter data to analyze the user reviews about drugs. We were able to retrieve twitter data successfully and focused only on drugs related tweets to identify their merits and demerits. We developed a classifier based on K-Nearest Neighbour classification algorithm and experiments were conducted on the twitter data.

## REFERENCES

[1] Xing Fang, Justin Zhan Sentiment analysis using product review data In: Journal of Big Data (2015) 2: 5. doi:10.1186/s40537-015-0015-2

[2] Hu M, Liu B (2004) Mining and summarizing customer reviews In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 168–177.. ACM, New York, NY, USA.

[3] Gann W-JK, Day J, Zhou S (2014) Twitter analytics for insider trading fraud detection system In: Proceedings of the second ASE international conference on Big Data.

[4] Lin, Jimmy, and Alek Kolcz. "Large-scale machine learning at twitter." In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, pp. 793-804. ACM, 2012.

[5] Bian, Jiang, Umit Topaloglu, and Fan Yu. "Towards large-scale twitter mining for drug-related adverse events." In Proceedings of the 2012 international workshop on Smart health and wellbeing, pp. 25-32. ACM, 2012.

[6] Johan Bollen, Bruno Gonçalves, Guangchen Ruan, Huina Mao Happiness is assortative in online social networks. Artificial Life volume 17 issue 3, 2011 237-251.

[7] Ramesh R, Divya G, Divya D, Merin K Kurian Big Data Sentimental Analysis using Hadoop. International Journal for Innovative Research in Science & Technology, Volume 1, Issue 11, August 2015.

[8] Rasheed M. Elawady, Sherif Barakat, Nora M.Elrashidy, "Different Feature Selection for Sentiment Classification, "International Journal of Information Science and Intelligent System, 3(1): 137-150, 2014.

[9] Mahmoud Elgamal , Sentiment Analysis Methodology of Twitter Data with an application on Hajj season International Journal of Engineering Research & Science [Vol-2, Issue-1, January- 2016

[10] Sahana D S, Shantala C P, Girish L, Automatic Drug Reaction Detection Using Sentimental Analysis, International Journal of Advanced Research in Computer Engineering & Technology, Volume 4, Issue 5, May 2015.

[11] Friedman C. Discovering novel adverse drug events using natural language processing and mining of the electronic health record. In: Proceedings of the 12th conference on artificial intelligence in medicine (AIME); 2009.p. 1-5.

[12] Denecke K, Deng Y Sentiment Analysis in medical settings: New Opportunities and Challenges, Artif Intell Med, 2015 May, 64(1), 12-27. Doi: 10.1016/ j.artmed.2015.03.006 Epub 2015 May.